ELSEVIER
Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT®
Journal of Pure and Applied Algebra 186 (2004) 1–19
JOURNAL OF PURE AND APPLIED ALGEBRA
www.elsevier.com/locate/jpaa

Every finite semigroup is embeddable in a finite relatively free semigroup

George M. Bergman

Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA

Received 25 August 2002; received in revised form 12 February 2003
Communicated by J. Rhodes

Abstract

The title result is proved by a Murskii-type embedding.

Results on some related questions are also obtained. For instance, it is shown that every finitely generated semigroup satisfying an identity $\xi^d = \xi^{2d}$ is embeddable in a relatively free semigroup satisfying such an identity, generally with a larger d; but that an uncountable semigroup may satisfy such an identity without being embeddable in any relatively free semigroup.

It follows from known results that every finite group is embeddable in a finite relatively free group. It is deduced from this and the proof of the title result that a finite monoid S is embeddable by a monoid homomorphism in a finite (or arbitrary) relatively free monoid if and only if its group of invertible elements is either $\{e\}$ or all of S.

MSC: Primary: 20M05; secondary: 20M07; 20M30

1. Introduction

The proof of the title result will be given (to use the jargon of computer programming) in "top-down" format: The next section gives the skeleton of the argument, the two that follow fill in the steps sketched, assuming a family of semigroup words given having certain properties, and finally, in Section 5, such a family of words is displayed and the required properties checked.

To help the reader keep track of the statements assumed at various points that are to be proved later, whenever we make such a statement we will display it with a label shown in the form "(n ↓)", and when the result has been verified, we will note this

E-mail address: gbergman@math.berkeley.edu (G.M. Bergman).

0022-4049/03/$ - see front matter
doi:10.1016/S0022-4049(03)00126-9

by writing "($n\surd$)". However, that verification may use other statements marked "$\downarrow$" which remain to be proved; thus, our proof will be complete only when all of our "$\downarrow$"s are "$\surd$"ed.

The title result answers a question of John Rhodes and Benjamin Steinberg (personal communication), and will be used in [8, Chapter 2]. The last two sections of this paper obtain some related results and note some further questions.

2. The framework of the proof

Let $S=\{a_1,\dots,a_n\}$ be a finite nonempty semigroup. For each $i,j\in\{1,\dots,n\}$, let $i*j\in\{1,\dots,n\}$ be the unique value such that

$$a_ia_j=a_{i*j}. \tag{1}$$

In Section 5 we shall define n distinct semigroup words in two indeterminates ξ and η,

$$A_i(\xi,\eta)\quad (i\in\{1,\dots,n\}). \tag{2 $\downarrow$}$$

Assuming these given, let $\mathbf{V}$ be the variety of semigroups defined by the n^2 identities

$$A_i(\xi,\eta)A_j(\xi,\eta)=A_{i*j}(\xi,\eta)\quad (i,j\in\{1,\dots,n\}). \tag{3}$$

If $F(x,y)$ is the relatively free semigroup on two generators x and y in $\mathbf{V}$, or, indeed, in any subvariety of $\mathbf{V}$, then the identities (3) guarantee that the map $S\to F(x,y)$ given by

$$a_i\mapsto A_i(x,y) \tag{4}$$

is a homomorphism. To complete the proof, we shall construct a semigroup T containing two elements x and y such that

$$T \text{ is finite,} \tag{5 $\downarrow$}$$

$$T \text{ satisfies the identities (3),} \tag{6 $\downarrow$}$$

$$\text{The elements } A_i(x,y)\ (i\in\{1,\dots,n\}) \text{ of } T \text{ are distinct.} \tag{7 $\downarrow$}$$

By (6), the variety generated by T is contained in $\mathbf{V}$. If we take $F(x,y)$ free in that variety, then (5) implies that $F(x,y)$ is finite, and (7) implies that the homomorphism (4) of S into $F(x,y)$ is an embedding, establishing the title result.

I am grateful to M. Volkov for pointing out to me that a similar technique for the construction and study of semigroup varieties was introduced in 1968 by Murskii [6].

In the next two sections we shall assume that words (2) are given, and satisfy various properties which we will state as they are needed. The reader may, of course, peek ahead to Section 5 and see what these words are, if and when he or she feels this would be helpful.

3. The construction of T

Assuming the family of semigroup words (2) given, let us take for T the semigroup presented (as a semigroup—without assuming (6)) by three generators, x, y and 0, and three families of relations: First, the particular cases of (6) gotten by substituting x for ξ and y for η:

$$A_i(x,y)A_j(x,y) = A_{i*j}(x,y) \quad (i,j \in \{1,\dots,n\}), \tag{8}$$

second, the five relations making 0 a zero element of T:

$$x0 = 0, \quad 0x = 0, \quad y0 = 0, \quad 0y = 0, \quad 00 = 0, \tag{9}$$

and finally, the infinite family of relations saying that

Every word in x and y which is *not* a subword of a product $A_{i_1}(x,y)\dots A_{i_r}(x,y)$ ($r \geqslant 1$, $i_1,\dots,i_r \in \{1,\dots,n\}$) is equal to 0 in T. (10)

(Throughout this note, a "subword" of a word will mean a string of *consecutive* symbols in that word.) In the presence of (9), the family of relations (10) is clearly equivalent to the smaller family of relations saying that

Every word in x and v which is *minimal* (under passing to subwords) for the property of not being a subword of a product $A_{i_1}(x,y)\dots A_{i_r}(x,y)$ is equal to 0 in T. (11)

Observe that the set of words in x, y and 0 having no subwords to which any of the reduction rules (8), (9) and (11) can be applied consists of 0, and all words $U(x,y)$ such that $U(x,y)$ is a subword of a product of the $A_i(x,y)$'s (by (10)), but has no subword which is a full product of two $A_i(x,y)$'s (by (8)). We now wish to prove that *distinct words in this set represent distinct elements of T.*

The conditions on a semigroup or similar algebraic object presented by generators and relations, where the relations are treated as "reduction rules", for such a conclusion to hold, have been known under various names, and stated with various degrees of precision; the formulation we will follow is that of [2]. Roughly speaking, it is proved there that what must be verified is, first, that no word admits an infinite sequence of successive reductions (applications of the reduction formulas, in our case (8), (9) and (11), to subwords), and, secondly, that for every minimal case of a word that can be reduced in two conflicting ways—that is, every case where either the left-hand side V of one reduction formula is a subword of the left-hand side W of another, or where some product of words UVW has the property that UV forms the left-hand side of one such formula and VW that of another—the results of reducing W, respectively UVW, in these two ways can subsequently be brought to equality by further application of reductions from our family. Following [2], we shall call such cases of words that can be reduced in two ways "ambiguities" of our reduction system, and call the confluence condition that must be verified "resolvability" of the ambiguity. (For details see Section 1 of that paper, which develops the result in the context of unital rings, Section 9.1 which notes the simplified form it takes in the case of monoids, there called "semigroups with 1", and Section 9.2, which notes that corresponding statements are valid for nonunital rings and semigroups without 1.)

In the present situation, it is immediate that no infinite sequence of reductions can be applied successively to any word, because each reduction decreases the length of the word.

To verify the resolvability of all ambiguities, we first note that none of our reductions turns a word containing the letter 0 into one not containing 0. It is easily deduced that given any ambiguity such that at least one of the two reductions involved is a case of (9), the word resulting from each reduction can be further reduced to 0, so such ambiguities are resolvable. We also see that for any ambiguity such that *both* reductions are instances of (11), both sides likewise reduce to 0.

So it remains to consider the cases where either both reductions are instances of (8), or one is an instance of (8) and the other an instance of (11). In verifying that these are resolvable, we will use the following property of the words A_i to be defined:

$$\begin{array}{l}\text{There are no inclusions or overlaps among the words } A_i(x,y).\\ \text{That is, none of these words is a subword of any other,}\\ \text{and there are no choices of words } U,\ V,\ W \text{ in } x \text{ and } y \text{ such that}\\ \text{each of the words } UV \text{ and } VW \text{ belongs to } \{A_i(x,y) \mid i \in \{1,\dots,n\}\}.\end{array} \qquad (12\downarrow)$$

From (12), it is easy to see that the only ambiguities in which both reductions are instances of (8) are those arising from products UVW, where $U = A_i(x,y)$, $V = A_j(x,y)$, $W = A_k(x,y)$ for some $i,j,k \in \{1,\dots,n\}$. In this situation, after we apply the two reductions in question to this product, one more application of (8) reduces the resulting words to $A_{(i*j)*k}(x,y)$ and $A_{i*(j*k)}(x,y)$ respectively, which are equal by associativity of S (cf. (1)).

In the case where one reduction is an instance of (8) and the other an instance of (11), it is not hard to deduce from (12) that either our minimal ambiguously reducible word has the form UVW with UV equal to the left-hand side of an instance of (8), and W not an initial segment of any product of $A_i(x,y)$'s, or we are in the mirror-image of this situation; by symmetry it suffices to consider the former case. Note that after one applies the reduction coming from (8), the resulting expression still involves a word $A_i(x,y)$ followed by a string that is not an initial segment of such a word. In view of (12), such an expression is not a subword of any product of $A_i(x,y)$'s, hence, by (11), it reduces to 0. Since the other reduction of UVW, by applying (11) to VW, gives $U0$ which reduces to 0, these ambiguities are also resolvable.

Since all ambiguities in our reduction system are resolvable, the semigroup presented using the relations (8), (9) and (11) has a normal form consisting of those words in x, y and 0 which are irreducible with respect to that reduction system; that is

$$\begin{array}{l}\text{Every element of } T \text{ is either } 0, \text{ or has a unique expression as a}\\ \text{word } U(x,y) \text{ in } x \text{ and } y \text{ which occurs as a subword of a product}\\ \text{of one or more words from the set } \{A_i(x,y) \mid i \in \{1,\dots,n\}\}, \text{ but}\\ \text{does not contain as subword a full product of two factors from that set.}\end{array} \qquad (13)$$

Since, in particular, every element $A_i(x, y)$ is of this form, the $A_i(x, y)$ represent distinct elements of T, proving (7$\surd$) (modulo statements which remain to be verified).

Note also that if L is the greatest of the lengths of the $A_i(x, y)$'s, then a word as in (13) can have length at most $3L - 2$. (It can consist of at most an $A_i(x, y)$ flanked on either side by a word one letter short of being an $A_i(x, y)$.) Hence there are only finitely many such words, giving (5$\surd$).

(For an estimate which better shows the order of magnitude of card(T), note that each nonzero word as in (13) can be obtained—possibly nonuniquely—by choosing a product of three $A_i(x, y)$'s, choosing some letter within the first of these three factors to be the first letter of U, and some letter in the whole word to be the last. This gives $< n^3 \cdot L \cdot 3L$ words in x and y. Also counting the word 0, we conclude that $\mathrm{card}(T) \leqslant 3n^3L^2$.)

The next section will be devoted to verification of (6), i.e., to showing that T belongs to $\mathbf{V}$.

4. *T* belongs to V

To prove this we must show that

$$\begin{aligned}&\text{For every pair of elements } X, Y \in T, \text{ we have}\\ &A_i(X, Y)A_j(X, Y) = A_{i*j}(X, Y) \text{ for all } i, j \in \{1, \ldots, n\}.\end{aligned} \tag{14 $\downarrow$}$$

In view of (10), we can expect that most choices of X and Y will cause both sides of the equations of (14) to reduce to 0 in T. In determining what the exceptions are and analyzing these, we will invoke several more properties of the words A_i.

To state the first of these, let a *run* of x's or y's within a word mean a block of consecutive x's or y's that is not contained in a larger such block. Then we shall assume

$$\begin{aligned}&\text{All words } A_i(\xi, \eta) \text{ have the same length } L > 6, \text{ and all begin}\\ &\text{with a run of more than } L/3 \text{ but fewer than } L/2\ \xi\text{'s, and}\\ &\text{end with a run or more than } L/3 \text{ but fewer than } L/2\ \eta\text{'s.}\end{aligned} \tag{15 $\downarrow$}$$

We can see from (15) that a subword W of a product of the words $A_i(x, y)$ will have no run of $\geqslant L/2$ x's or y's; but will, within every interval of length L, have a run of $> L/3$ x's or $> L/3$ y's; and moreover that the run of x's, if any, following every run of $> L/3$ y's will also have length $> L/3$, unless it is terminated by the end of our word W, in which case it may have shorter length, and similarly that the run of y's, if any, preceding every run of $> L/3$ x's will have length $> L/3$ unless cut off by the beginning of W. Finally, the points of transition between runs of $> L/3$ y's and the following runs of $> L/3$ x's (where one of these two runs may have smaller length if cut off by the beginning or end of the word) will be spaced at intervals of length precisely L, and the subword between each such transition and the next will be one of the words $A_i(x, y)$.

Now suppose W is a word in x and y of length >1, which is not only a subword of a product of the $A_i(x,y)$, but has the property that for some $r>L/3$, W^r is still a subword of such a product. (We are, of course, looking for words which can be used as X and/or Y in (14) without causing both sides to reduce to 0. The situation where X or Y or both have length 1, though simpler than the situation where they both have greater length, will be postponed to the end of this section to allow us to use there some special cases of observations developed for this harder situation.) We see, first of all, that W must involve both the letters x and y, since otherwise W^r would constitute a run of $>2L/3$ x's or y's. Also, the fact that W^r has length $>2L/3$, together with (15), implies that it must contain a run of $>L/6$ x's or y's, which, as $L>6$, eliminates the possibilities $W=xy,\ yx$; so W must have length at least 3. This gives W^r length $>L$, hence it must contain a run of $>L/3$ x's or y's, along with the complementary letter that terminates each end of that run. Using our observation on the regular spacing of transitions between runs of y's and x's of that length, it is now not hard to deduce

> If W is a word in x and y of length >1, such that for some $r>L/3$, the word W^r is a subword of a product of the words $A_i(x,y)$, then W has length sL for some positive integer s, and can be obtained from a product of s of the words $A_i(x,y)$ by moving some (possibly empty) initial subword of length $\leqslant L/2$ from the beginning to the end, or some (possibly empty) final subword of length $\leqslant L/2$ from the end to the beginning. (16)

Let us note further that if W_1 and W_2 are words both having the form described in (16), and if moreover both $W_1^2W_2$ and $W_1W_2^2$ are subwords of products of the $A_i(x,y)$, then using again the periodicity of transitions from long runs of y's to long runs of x's, we see that the lengths of the transposed subwords in the descriptions of W_1 and W_2 as in (16), and the directions in which they are transposed, must be the same (assuming for the moment that if a transposed segment has length exactly $L/2$ it is transposed from the beginning to the end, to make the description in (16) unique). In fact, these transposed subwords are forced to be identical if we assume yet another property of the $A_i(x,y)$:

> For $i\neq j$, the words $A_i(x,y)$ and $A_j(x,y)$ differ both in their initial subwords of length $[L/2]$ and in their final subwords of length $[L/2]$, where $[L/2]$ denotes the greatest integer $\leqslant L/2$. (17 ↓)

That is, each half of one of the A_i determines the other half. It follows that one cannot glue two *different* words of some length $L_0\leqslant L/2$ onto the same side of the same word of length $L-L_0$ and get in each case a word in $\{A_j(x,y)\mid i\in\{1,\ldots,n\}\}$. The assertion preceding (17) is now clear, giving us

> If X and Y are words of length >1 such that for some $i\in\{1,\ldots,n\}$, $A_i(X,Y)$ does not reduce to 0 in T, then either there exist

$i_1,\ldots,i_s,j_1,\ldots,j_t \in \{1,\ldots,n\}$ $(s,t \geqslant 1)$ and factorizations $A_{i_1} = BA'_{i_1}$ and and $A_{j_1} = BA'_{j_1}$ with B of length $\leqslant L/2$, *possibly empty*, such that $X = A'_{i_1}A_{i_2}\ldots A_{i_s}B$, $\quad Y = A'_{j_1}A_{j_2}\ldots A_{j_t}B$, or the situation is the right−left mirror image of this one. (18)

Note: In (18) above we explicitly allow B to be the empty string. Except where such an explicit exception is made, all letters appearing in equations in T are understood to denote elements of T. Note also that for brevity we have written A_{i_1}, etc., instead of $A_{i_1}(x,y)$, etc. in (18). We shall do the same from time to time without comment in the remainder of this section.

The completion of our proof of (14) in the case where X and Y have length > 1 requires two more properties of the A_i. The first seems strong, but will be easy to build into our construction of the A_i using the finiteness of the semigroup S:

For all $i,j_1,j_2 \in \{1,\ldots,n\}$, the element $A_i(a_{j_1},a_{j_2}) \in S$ is an idempotent $a_h \in S$, which is independent of i (but in general depends on j_1 and j_2). (19 ↓)

Since the $A_i(x,y) \in T$ satisfy the same relations as the $a_i \in S$, it follows from (19) that in the case of (18) where the word B is empty, the equation from (14) that we need to verify reduces to one with a certain element $A_h(x,y)$ on the right-hand side, and the square of that element on the left. Moreover that element is idempotent, so the equation holds.

To handle the case where B may not be empty, let us simplify the notation of (18) by writing $B^{-1}W$ for the result of removing an initial string B from a word W, assuming W begins with that string; the symbol will be undefined otherwise. (In using this notation, we must keep in mind that if evaluating W in T gives the same result as evaluating another word W', this may not be true of $B^{-1}W$ and $B^{-1}W'$.) Then the expressions for X and Y given in (18) (ignoring, without loss of generality, the mirror-image case) take the forms $X = B^{-1}A_{i_1}A_{i_2}\ldots A_{i_s}B$ and $Y = B^{-1}A_{j_1}A_{j_2}\ldots A_{j_t}B$. When we substitute these into the two sides of (14) we get—prior to reduction—the same expressions we got when B was empty, except for a B^{-1} on the left, and a B on the right. However, the B^{-1} on the left may prevent us from calculating in T as we did before. To get around this, we shall call on one more property of the A_i.

It follows easily from the finiteness of the semigroup S that there exists a positive integer d such that

$$a_i^{2d} = a_i^d \quad (i \in \{1,\ldots,n\}). \tag{20}$$

(If S were a group, d would be called an exponent of that group.) Recalling that L is the common length of the words $A_i(\xi,\eta)$, which all begin with $> L/3$ ξ's, we shall assume that

$$L/3 \geqslant d, \tag{21 ↓}$$

so that each $A_i(\xi,\eta)$ begins with $>d$ ξ's. It follows that each of the expressions we want to prove equal begins with

$$B^{-1}(A_{i_1}A_{i_2}\dots A_{i_s})^{d+1} = B^{-1}(A_{i_1}A_{i_2}\dots A_{i_s})(A_{i_1}A_{i_2}\dots A_{i_s})^{d}$$

which by (20) is equal in T to

$$B^{-1}(A_{i_1}A_{i_2}\dots A_{i_s})(A_{i_1}A_{i_2}\dots A_{i_s})^{2d} = B^{-1}(A_{i_1}A_{i_2}\dots A_{i_s})^{d}(A_{i_1}A_{i_2}\dots A_{i_s})^{d+1}.$$

Thus, the product $(A_{i_1}A_{i_2}\dots A_{i_s})^{d+1}$ at the left-hand edge of the expression we want to reduce, which was originally "marred" by the B^{-1}, now appears as a genuine factor, insulated from the B^{-1} by the word $(A_{i_1}A_{i_2}\dots A_{i_s})^{d}$ which has length larger than that of B, since $Ld > L/2$. Hence we can now perform the same calculations we did in the case where B was empty, ending up with the same equality, except for an extra factor of $B^{-1}(A_{i_1}A_{i_2}\dots A_{i_s})^{d}$ on the left of each side, and an extra factor of B on the right, which do not disturb the equality.

This completes the proof that if X and Y are words in x and y, both of length >1, which when substituted into at least one of the words A_i do not give 0, then when substituted into the identities (3), they give equality. Of course, if X and Y are words of arbitrary lengths which do give 0 when substituted into all A_i, then these cases of (3) reduce to "$0=0$", and this includes the case where one or both of X and Y is not purely a word in x and y, but involves 0. So we have proved all cases of (14) except those where at least one of X and Y has length 1, and where the equation to be proved does not reduce to $0=0$.

Let us start with the case where one of X or Y, without loss of generality X, still has length >1, while Y has length 1. Then (16) describes the form of X. Calling on a final assumption about of our words A_i,

All runs of ξ's and all runs of η's in every $A_i(\xi,\eta)$ have length >1, (22 $\downarrow$)

we see that when we substitute X and Y into any A_i, the result has a subword of the form $X^2Y^mX^2$ with $1<m<L$. But observe that each factor X^2 has a point of transition between a run of $>L/3$ y's and a run of $>L/3$ x's, and if there were no Y^m between them, the distance between any such transition point in one factor X^2 and any such transition point in the other would be a multiple of L; hence with the length-m factor Y^m inserted between them, the distance is not a multiple of L, contradicting the properties we noted for subwords of products of A_i. So this case is excluded.

Assuming X and Y both have length 1, note that if they were the same letter, then any $A_i(X,Y)$ would be a run of that letter of length L; but we know that no runs of length $\geqslant L/2$ occur. Also, if X were y and Y were x, then $A_i(X,Y)$ would begin with a run of $>L/3$ y's *not* followed by a run of $>L/3$ x's, which is again impossible. This leaves us with the case $X=x$, $Y=y$. In that case, of course, the desired relations hold by (8).

This completes the proof of (14$\surd$), which was a restatement of (6$\surd$).

5. The words A_i

Since for any value of d satisfying (20), the same equation is also satisfied by all multiples of d, we can assume the d of (20) chosen so that

$$d > 1. \tag{23}$$

We can now give the formula for the A_i's promised in ($2\surd$):

$$A_i(\xi, \eta) = \xi^{d^2(n+1)}(\xi^{di}\eta^{d(n+1-i)})^d \eta^{d^2(n+1)}. \tag{24}$$

Note that each of the three factors comprising the right-hand side of (24) has length $d^2(n+1)$, so $L = 3d^2(n+1)$. Conditions ($15\surd$), ($17\surd$), ($21\surd$) and ($22\surd$) are immediate from the above expression.

It is also not hard to check (12): The large runs of ξ and η at the beginnings and ends of the words (24) insure that the only kind of overlap that could occur between A_i and A_j would have the initial run of ξ's in one word containing that in the other, and ending at the same point. But then the lengths of the following runs of η's would have to be the same, forcing i to equal j and the overlap to be equality, establishing ($12\surd$).

Finally, to get (19), note that by (20) all dth powers in S are idempotent, hence that for every $r > 0$, a (dr)th power can be simplified to the corresponding dth power. Hence $A_i(a_{j_1}, a_{j_2})$ simplifies to $a_{j_1}^d(a_{j_1}^d a_{j_2}^d)^d a_{j_2}^d$. This is independent of i; moreover, the middle factor is left divisible by the idempotent element $a_{j_1}^d$ and so can absorb the $a_{j_1}^d$ on the left, and can similarly absorb the $a_{j_2}^d$ on the right, so the expression simplifies further to $(a_{j_1}^d a_{j_2}^d)^d$, which, being a dth power, is itself idempotent. This completes the proof of ($19\surd$). Thus, by the argument outlined in Section 2, we have proved

Theorem 1. *Every finite semigroup S is embeddable in a finite relatively free semigroup on two generators.* □

We remark that the above proof embeds S in a free semigroup in a variety larger than that generated by S. The following example of Rhodes (personal communication) shows that we cannot in general use a free semigroup in the variety generated by S itself. Let S be the semigroup $\{a, b, ab, 0\}$ where all products except $a \cdot b = ab$ equal 0. It is easy to check that in the variety generated by S, the free semigroup on a nonempty set G consists of the members of G, the pairwise products of distinct members of G (counting order), and a zero element 0, and that all products except products of distinct generators give 0. These free semigroups have no pairs of elements whose product is zero in one order but not in the other, hence S does not embed in such a semigroup.

(On the other hand, this S can be embedded in a relatively free semigroup in a variety much less elaborate than that of our proof. E.g., in the variety defined by the identities $\xi^2\eta = \xi^2 = \eta\xi^2$, saying that every square is a zero element, one can embed S in the free object on x and y by sending a to x and b to yx.)

6. Results and counterexamples for infinite semigroups

If we want a semigroup S to be embeddable in a finite relatively free semigroup F, we clearly cannot assume less than that S is finite; so in that sense Theorem 1 is best possible. But what if we delete the requirement that F be finite?

One strong restriction on embeddability is noted in

Proposition 2. *Let S be a semigroup admitting no homomorphism into the additive semigroup of positive integers* (*e.g.*, *any semigroup containing an idempotent element*, *or more generally*, *having a solution to $xy=x$*, *or $xy=y$*). *Then if S is embeddable in a relatively free semigroup*, *there exists an integer d such that S satisfies the identity $\xi^d=\xi^{2d}$.*

Proof. Suppose S is embeddable in a free semigroup F in the variety $\mathbf{V}$. Now F admits a homomorphism to the free semigroup on one generator x in $\mathbf{V}$, hence by assumption, that relatively free semigroup is not isomorphic to the additive semigroup of positive integers. This means that an equality $x^m=x^n$ $(m<n)$ holds in that free semigroup, from which one can deduce an equality $x^d=x^{2d}$, which is thus an identity of $\mathbf{V}$, and hence of S. □

Surprisingly, the converse to the last sentence of the above proposition holds for finitely generated semigroups, as we shall now prove by a modification of the method of Theorem 1. To do this, we must replace the systems of identities (3) and relations (8) used in that proof, which were based on the formulas (1) for computing in S, with something more general. To formulate the generalized result, consider any finitely generated semigroup S with generating set $a_1,\dots,a_n$, and for each $a\in S$, define the *reduced expression* for a to mean the expression for a as a product of these generators which is of least length, and, among all expressions of that length, is lexicographically first. We will call any word in the symbols $a_1,\dots,a_n$ *reduced* if it is the reduced expression for the element of S it represents. Now

> Let $\mathrm{Red}(S;a_1,\dots,a_n)$ denote the set of all ordered pairs (P,Q) such that $P=P(a_1,\dots,a_n)$ is a minimal word in $a_1,\dots,a_n$ which is *not* reduced (i.e., a word which is not reduced, but all of whose proper subwords are), and $Q=Q(a_1,\dots,a_n)$ is the reduced word representing the same element. (25)

As noted in [2, Section 5.3] (for algebras over a field; but again, the case of semigroups holds for the same reasons), the set of relations $P(a_1,\dots,a_n)=Q(a_1,\dots,a_n)$ where $(P,Q)\in\mathrm{Red}(S;a_1,\dots,a_n)$ will constitute a system of reduction formulas presenting the semigroup S, which will be terminating (have the property that no word admits an infinite sequence of successive reductions), and all of whose ambiguities will be resolvable.

We summarize in the following lemma some steps in our proof of Theorem 1 that carry over to this general situation with almost no change. (Note that we have not yet assumed an identity $\xi^d=\xi^{2d}$.)

Lemma 3. *Let S be a semigroup generated by $n > 0$ elements $a_1, \ldots, a_n$, and let $A_1(\xi,\eta), \ldots, A_n(\xi,\eta)$ be n distinct semigroup words in two indeterminates, satisfying* (12), (15), (17), *and* (22). *Let T be the semigroup presented by three generators x, y and* 0, *relations* (9) *used earlier, the relations* (11) *determined by the words $A_1(\xi,\eta), \ldots, A_n(\xi,\eta)$, and, in place of the n^2 relations* (8), *the* (*possibly infinite*) *family of relations*

$$P(A_1(x,y), \ldots, A_n(x,y)) = Q(A_1(x,y), \ldots, A_n(x,y)), \tag{26}$$

where (P,Q) ranges over the set $\mathrm{Red}(S; a_1, \ldots, a_n)$ *defined in* (25) *above.*

Then (i) *Each element of T can be represented uniquely as* 0 *or as a word in x and y which is a subword of a product of the words $A_i(x,y)$, but does not contain as a subword the left-hand side of any relations* (26).

(ii) *The map $a_i \mapsto A_i(x,y)$ is a semigroup embedding of S in T.*

(iii) *If W is a word in x and y of length > 1 such that for some $r > L/3$, W^r represents a nonzero element of T, then W has the form $B^{-1}A_{i_1}A_{i_2}\ldots A_{i_s}B$* (*notation as in Section* 4) *or $BA_{i_1}A_{i_2}\ldots A_{i_s}B^{-1}$* (*mirror-image of that notation*) *for some possibly empty word B of length $\leqslant L/2$, and some $i_1, \ldots, i_s \in \{1, \ldots, n\}$* ($s \geqslant 1$).

(iv) *If X and Y are elements of T such that for some $i \in \{1, \ldots, n\}$, $A_i(X,Y) \neq 0$, then either $X = x$ and $Y = y$, or $X = B^{-1}A_{i_1}A_{i_2}\ldots A_{i_s}B$ and $Y = B^{-1}A_{j_1}A_{j_2}\ldots A_{j_t}B$ for some possibly empty word B of length $\leqslant L/2$ and some $i_1, \ldots, i_s$, $j_1, \ldots, j_t \in \{1, \ldots, n\}$* ($s, t \geqslant 1$); *or the situation is the left–right mirror image of this.*

Method of Proof. Like the corresponding steps in the proof of Theorem 1, with the following modifications in the verification of (i): Where the resolvability of an ambiguity based on the left-hand sides of two equations from (8) was previously obtained from the associativity of S, the corresponding statement involving left-hand sides of two equations from (26) follows from the resolvability of all ambiguities in the reduction system $\mathrm{Red}(S; a_1, \ldots, a_n)$; and where previously, the termination of the reduction procedure followed from the fact that the reductions were length-decreasing, one now calls on the fact that such reductions *either* decrease the length of a word, *or* preserve the length and reduce the lexicographic position of the string of indices $i_1, \ldots, i_s$ associated with word's longest subword of the form $A_{i_1}A_{i_2}\ldots A_{i_s}$. □

We can now obtain the promised partial converse to Proposition 2.

Theorem 4. *Every finitely generated semigroup S satisfying the identity*

$$\xi^d = \xi^{2d} \tag{27}$$

for some positive integer d can be embedded in a relatively free semigroup on two generators satisfying the identity $\xi^{d'} = \xi^{2d'}$ for some positive integer d'.

Sketch of Proof. Let $a_1, \ldots, a_n$ be a generating set for S, let $A_1(\xi,\eta), \ldots, A_n(\xi,\eta)$ be the same words (24) that we used in the proof of Theorem 1, with the d in their definition taken to be the value in (27), which, as before, we increase if necessary so that (23) holds, and let T be the semigroup constructed from these data as in Lemma 3 above. We claim that T satisfies the identities

$$P(A_1(\xi,\eta),\dots,A_n(\xi,\eta)) = Q(A_1(\xi,\eta),\dots,A_n(\xi,\eta))$$
$$\text{for all } (P,Q)\in \mathrm{Red}(S;a_1,\dots,a_n). \tag{28}$$

Lemma 3(iv) shows that the words X, Y in our generators which, when substituted for ξ and η in (28), yield a possibly nontrivial relation to be checked are as in the proof of Theorem 1. Identity (27), corresponding to our earlier condition (20), again yields (19) (cf. paragraph before Theorem 1), and (19) and (27) again reduce the verification of the hard case of the relations we must verify to the equality between two powers of a common idempotent element. In the remaining case, namely $X=x$, $Y=y$, the relations are assured as before by our presentation (26) of T. As in the proof of Theorem 1, it follows that the map (4) of S into the free semigroup F on two generators in the variety generated by T is a one-to-one homomorphism.

It remains to show that this variety satisfies an identity $\xi^{d'}=\xi^{2d'}$. A quick way is to note that, assuming S nonempty, the identity $\xi^d=\xi^{2d}$ for that semigroup shows that it cannot be mapped homomorphically into $1+\mathbb{N}$, hence neither can the relatively free semigroup F in which we have embedded it, hence by Proposition 2, F satisfies an identity of the same sort.

Alternatively, one can show directly that T, and hence the variety **V** it generates, satisfies the identity $\xi^L=\xi^{2L}$, where as in Section 5, $L=3d^2(n+1)$, by using Lemma 3(iii) to restrict the words W on which we must test this identity, and making use of the identity $\xi^d=\xi^{2d}$ satisfied by S. $\square$

Can a similar method be applied to nonfinitely generated semigroups? Given a *countable* semigroup $S=\{a_1,a_2,\dots,a_i,\dots\}$ satisfying an identity $\xi^d=\xi^{2d}$, it seems plausible that one may be able to choose words A_i $(i=1,2,\dots)$ satisfying some of the key conditions we have used above, say (12), (17) and (22), and apply the same general idea to embed S in a semigroup with the desired properties. However, we would have to make some major adjustments in our arguments. Certainly the infinite family of words A_i could not satisfy (15), i.e., all have the same length. In fact, they could not be obtained from any "closed form" expression using exponents as the parameters to be varied if the semigroup T we are constructing is to satisfy an identity $\xi^{d'}=\xi^{2d'}$, since application of this identity would kill the distinctions among all but finitely many such words. So a more sophisticated coding technique would be needed.

Another approach to proving Theorem 4 with "finitely generated" replaced by "countable" might be to try to embed an arbitrary countable semigroup satisfying an identity $\xi^d=\xi^{2d}$ in a finitely generated semigroup satisfying an identity $\xi^{d'}=\xi^{2d'}$.

I leave these possibilities for others to explore.

If we want to embed *uncountable* semigroups in relatively free semigroups, even the above vague ideas obviously would not work. There is, in fact, a nontrivial obstruction to such embeddings.

Lemma 5. *If S is a semigroup, then the following conditions are equivalent*:

(i) *S has uncountably many isomorphism classes of finitely generated subsemigroups.*
(ii) *For some positive integer n, the set of congruences on the free semigroup F on n generators induced by homomorphisms into S is uncountable.*

If these equivalent condition hold, then S cannot be embedded in any relatively free semigroup.

(*More generally, the corresponding facts are true for algebras of any type that involves at most countably many operations and where all operations have finite arities.*)

Proof. Assuming (i), there must be some integer n such that S has a set of uncountably many pairwise nonisomorphic n-generated subsemigroups. Choose a family of n generators for each of these. Each such generating family determines a homomorphism from the free semigroup F on n generators into S, and these homomorphisms induce distinct congruences on F, establishing (ii).

To get the converse, note that any finitely generated semigroup admits only countably many n-element generating families, hence corresponds to at most countably many congruences on the free n-element semigroup. Hence if, as in (ii), n-generator subsemigroups of S induce uncountably many such congruences, there must be uncountably many isomorphism classes of such subsemigroups, giving (i).

Now if S can be embedded in a semigroup F free in a variety $\mathbf{V}$, then each finitely generated subsemigroup of S embeds in a finitely generated subsemigroup of F, which will be contained in a subsemigroup $F_m \subseteq F$ free on finitely many generators. So we have embeddings of all the finitely generated subsemigroups of S in countably many relatively free semigroups, each of which, being countable, has at most countably many finitely generated subsemigroups. Hence an S admitting such an embedding cannot satisfy (i).

To see the parenthetical generalization, note that the assumption of at most countably many operations, all of finite arities, still guarantees that finitely generated algebras are countable. (We must still restrict attention to *finitely* generated subalgebras of S, to be sure that a countable algebra has only countably many such subalgebras.) □

From this we can get

Corollary 6. *There exists a semigroup S satisfying the identity $\xi^2 = \xi^3$* (*and hence the identity $\xi^2 = \xi^4$*), *but not embeddable in any relatively free semigroup.*

Proof. It is shown in [5] that there exists an infinite 3-generator semigroup in which all squares are equal to a zero element, and hence which satisfies the identity $\xi^2 = \xi^3$. Adjoining a neutral element 1 we get a monoid, which we shall denote S_0, which clearly satisfies this identity. Let S_1 be the semigroup of all *partial set-maps* $S_0 \to S_0$, i.e., functions a from some subset $\mathrm{dom}(a) \subseteq S_0$ into S_0, with composition defined in the natural way, i.e., so that $ab(x)$ equals $a(b(x))$ if the latter is defined, and is undefined otherwise.

Let us name two sorts of elements of S_1: For each $x \in S_0$. let t_x denote the everywhere-defined function of left translation by x, and for each element $x \in S_0$ and subset $P \subseteq S_0$, let $c_{x,P}$ denote the "collapsing" function having domain P, and sending all members of P to x. These are all distinct, except that $c_{x,\varnothing}$ does not depend on x. Let us denote the latter element (the empty function) by $0 \in S_1$, but understand it to be counted in statements we make about elements $c_{x,P}$.

It is immediate that the elements of the forms t_x and $c_{x,P}$ comprise a subsemigroup $S \subseteq S_1$. The elements t_x form a subsemigroup isomorphic to S_0, hence, by choice of S_0, satisfying the identity $\xi^2 = \xi^3$. On the other hand, an element $c_{x,P}$ satisfies $c_{x,P}c_{x,P} = c_{x,P}$ if $x \in P$, $c_{x,P}c_{x,P} = 0$ otherwise. Hence if we substitute $c_{x,P}$ for ξ in $\xi^2 = \xi^3$, either both sides give $c_{x,P}$, or both sides give 0; so $\xi^2 = \xi^3$ is an identity of S.

Let us now show that S satisfies condition (ii) of Lemma 5, making an embedding in a relatively free semigroup impossible. Let x, y, z generate the semigroup $S_0 - \{1\}$, and for each nonempty subset $P \subseteq S_0 - \{1\}$ consider the semigroup relations satisfied by the four elements $t_x, t_y, t_z, c_{1,P} \in S$. For any word W in three semigroup variables, we see that $c_{1,P}W(t_x, t_y, t_z)c_{1,P}$ will equal $c_{1,P}$ if $W(x, y, z) \in P$, and 0 otherwise. Hence for distinct choices of P we get distinct sets of semigroup relations holding among these four elements. Since $S_0 - \{1\}$ is infinite, there are uncountably many choices for P, giving condition (ii) of the preceding lemma. □

I am indebted to the referee for pointing me to Morse and Hedlund's result [5] used above. (My original construction applied the same method to an infinite Burnside group, yielding an identity $\xi^d = \xi^{2d}$, with large d, and using the deep result of [1] rather than the elementary result of [5].)

Note that the uncountably many nonisomorphic 3-generator subsemigroups displayed in the above proof are each embeddable in a relatively free semigroup in a variety satisfying an identity $\xi^d = \xi^{2d}$, by Proposition 2. These embeddings require uncountably many such varieties (hence not all of them finitely based); the existence of uncountably many semigroup varieties was apparently first shown in [3].

If it should prove true that all countable semigroups satisfying identities $\xi^d = \xi^{2d}$ are embeddable in relatively free semigroups, one could ask whether Lemma 5 gives the only obstruction to such embeddings in the uncountable case. But this seems implausibly strong; it might, rather, be worth looking for other restrictions "in the spirit of" that lemma.

Let us now consider the other class of semigroups which Proposition 2 allows as potentially embeddable in relatively free semigroups—those admitting a homomorphism into the additive semigroup of positive integers, which we shall denote $1 + \mathbb{N}$.

Not every finitely generated semigroup in this class is embeddable in a relatively free semigroup. For instance, the semigroup presented by three generators and one commutativity relation, $S = \langle x, y, z \mid zy = yz \rangle$, admits a homomorphism to $1 + \mathbb{N}$ sending x, y and z to 1; but since the subsemigroup generated by x and y is absolutely free, S satisfies no nontrivial semigroup identities, so the only variety where it could possibly embed in a free semigroup is the variety of all semigroups. However, free semigroups in that variety have the property that any two elements which commute have a common power, which $y, z \in S$ do not. It is conceivable that every finitely generated semigroup satisfying a proper semigroup identity might admit such an embedding; but again, this seems very strong.

A test case I looked at was the class of semigroups gotten by taking a semigroup satisfying an identity $\xi^d = \xi^{2d}$, forming its direct product with $1 + \mathbb{N}$, and taking a finitely generated subsemigroup of that product. It turned out that these are indeed embeddable in relatively free semigroups. Generalizing the proof led to the following result (which is easily seen to include that class of examples). In the statement, note that since S is written multiplicatively and $1 + \mathbb{N}$ additively, w takes "products" to "sums".

Theorem 7. *Let S be a finitely generated semigroup which admits a homomorphism $w : S \to 1 + \mathbb{N}$, and which for some positive integer d satisfies the identity*

$$\xi^d \eta^d \xi^{2d} = \xi^{2d} \eta^d \xi^d. \tag{29}$$

Then S can be embedded in a relatively free semigroup on two generators which likewise satisfies an identity $\xi^{d'} \eta^{d'} \xi^{2d'} = \xi^{2d'} \eta^{d'} \xi^{d'}$.

Sketch of Proof. Let us begin by reducing to the case where S is generated by elements a satisfying $w(a) = 1$.

If S is generated by elements $a_1, \ldots, a_n$ not necessarily satisfying this condition, let us form the monoid $S \cup \{1\}$, extend w to the monoid homomorphism $S \cup \{1\} \to \mathbb{N}$ taking 1 to 0, form the product monoid $\mathbb{Z} \times (S \cup \{1\})$, and map this into $\mathbb{Z}$ by the homomorphism $w' : (m, a) \mapsto m + w(a)$. Clearly this product monoid still satisfies the identity $\xi^d \eta^d \xi^{2d} = \xi^{2d} \eta^d \xi^d$. Moreover, the $n + 1$ elements $(1 - w(a_1), a_1), \ldots, (1 - w(a_n), a_n)$ and $(1, 1)$ are each sent by w' to $1 \in \mathbb{Z}$, hence w' restricts to a semigroup homomorphism from the subsemigroup that these elements generate into $1 + \mathbb{N}$. That subsemigroup contains the elements

$$(1, 1)^{w(a_i) - 1}(1 - w(a_i), a_i) = (0, a_i) \quad (i \in \{1, \ldots, n\}),$$

which generate $\{0\} \times S \cong S$. Thus we have embedded S in a semigroup admitting a homomorphism to $1 + \mathbb{N}$ which carries all members of a finite generating set to 1.

So let us assume for the remainder of this proof that S is generated by elements $a_1, \ldots, a_n$ satisfying $w(a_i) = 1$, and let us form the system of reduction formulas $\mathrm{Red}(S; a_1, \ldots, a_n)$ as in (25), using this generating set. Note that in the present case, for each element (P, Q) of this set, P and Q will have the same length, namely the value of w on the element of S they both represent.

To take advantage of this homogeneity, let us choose our words $A_i(\xi, \eta)$ $(i \in \{1, \ldots, n\})$ all to have the same degree in each variable. We can achieve this by slightly modifying the words given in (24), and letting

$$A_i(\xi, \eta) = \xi^{2d^2(n+1)} (\xi^{di} \eta^{d(n+1-i)} \xi^{d(n+1-i)} \eta^{di})^d \eta^{2d^2(n+1)}. \tag{30}$$

These words clearly still satisfy (12), (15), (17), (21) and (22), but also have the property that (again writing L for their common length) they each have degree $L/2$ in each variable.

As before, Lemma 3 gives us a semigroup T in which S embeds, and we need to show that T satisfies the identities (28). Again, part (iv) of that lemma gives us a very limited class of values of X and Y to be checked. In checking these case, we previously used the fact that the various A_i had the same form, differing only in the powers of ξ^d and η^d occurring. The words (30) have this property, but we now also use the fact that, when we compare two of these words A_i and A_j, a higher power of ξ^d or η^d in one position is always balanced by a correspondingly lower power of the same factor in another position. These facts and the identity (29) allow us to rearrange the factors of any expression $A_i(X,Y)$ where X and Y have length greater than 1 to get any other such expression $A_j(X,Y)$. Since we have noted that for $(P,Q)\in \mathrm{Red}(S;a_1,\ldots,a_n)$, P and Q have the same length, the desired equalities follow. We similarly find that T will satisfy the identity of the form (29), but with $d'=L$ in place of d. □

This time, if we remove the finite generation assumption there are obstructions to such embeddings that affect even the countable case:

Lemma 8. *If a semigroup S admitting a homomorphism to $1+\mathbb{N}$ is embeddable in a relatively free semigroup, then the intersection of all congruences on S induced by homomorphisms into $1+\mathbb{N}$ has finite congruence classes.*

Proof. It clearly suffices to show that in any variety not satisfying an identity $\xi^d=\xi^{2d}$, the free semigroup $F(G)$ on any set G has the asserted property. Now if we define an equivalence relation on semigroup words in elements of G by calling two words U and V equivalent if each member of G occurs with the same multiplicity in U and in V, the equivalence classes are finite, and words in *different* equivalence classes can clearly be separated by homomorphisms $F(G)\to 1+\mathbb{N}$. The assertion follows. □

An example of a semigroup which admits a homomorphism to $1+\mathbb{N}$, is commutative, and hence satisfies all identities (29), and is countable, but which the above result shows is not embeddable in any relatively free semigroup, is $(1+\mathbb{N})\times\mathbb{Z}$. Indeed, it is easy to see that every homomorphism from this semigroup to $\mathbb{Z}$ has the form $(a,b)\mapsto pa+qb$ for some integers p and q, and that this map will be $(1+\mathbb{N})$-valued if and only if p is positive and $q=0$. All the homomorphisms to $1+\mathbb{N}$ so obtained induce the same congruence, namely $(a,b)\sim(a',b')\Leftrightarrow a=a'$, and this has infinite congruence classes $\{a\}\times\mathbb{Z}$.

One can refine the proof of Lemma 8 to get further restrictions. For instance, if S is a semigroup satisfying the hypothesis of that lemma, and for each $a\in S$ we let $c(a)$ denote the number of distinct elements $b\in S$ such that for all homomorphisms $w:S\to 1+\mathbb{N}$ one has $w(b)=w(a)$, then since the number of semigroup words of length m in a fixed finite set of variables grows exponentially in m, for each element $a\in S$, the integer-valued function $d\mapsto c(a^d)$ can grow at most exponentially in d. So if $f:\mathbb{N} \to \mathbb{N}$ is a function with faster than exponential growth, and satisfies $f(m_1+m_2) \geqslant f(m_1)+f(m_2)$ for all $m_1,\ m_2 \in \mathbb{N}$, so that $\{(a,b) \in (1+\mathbb{N}) \times \mathbb{Z} \mid |b| \leqslant f(a)\}$ is a subsemigroup S of $(1+\mathbb{N}) \times \mathbb{Z}$, then that subsemigroup, though it satisfies the conclusion of Lemma 8, is not embeddable in a relatively free semigroup.

7. Groups and monoids

Can we prove results analogous to Theorem 1 for classes of algebraic objects other than semigroups? The corresponding statement for groups is true, but for different reasons. We shall obtain it from

Lemma 9 (see Magidin [4, Lemma 2.5]; cf. Oates [7, Lemma 3.2]). *If G is a finite simple group which can be generated by n elements, then G is isomorphic to a direct factor in the free group on n generators in the variety generated by G.* □

M. Sapir and M.V. Volkov have both pointed out to J.Rhodes (personal communications) that since every finite group is embeddable in an alternating group on $\geqslant 5$ letters, which is simple and generated by two elements, the above fact implies

Theorem 10. *Every finite group is embeddable in a finite relatively free group generated by two elements.* □

However, as with semigroups, a finite nonsimple group need not be embeddable in a free group in the variety it generates. Here is an example suggested by A. Magidin (personal communication). Let G be the group of strictly upper triangular invertible matrices over the field $\mathbb{Z}/2\mathbb{Z}$,

$$G = \{I + ae_{12} + be_{23} + ce_{13} \mid a,b,c \in \mathbb{Z}/2\mathbb{Z}\}$$

and let $\mathbf{V}$ be the variety of groups generated by G. It is easy to see that multiplication in G is additive on the coefficients of e_{12} and e_{23}, hence that for any group word W in any number of variables, if each variable has even exponent-sum in W, then applied to elements of G, W takes values of the form $I + ce_{13}$. Since such elements have exponent 2, we see that for such W, $W^2 = 1$ is an identity of $\mathbf{V}$. On the other hand, G has elements of order 4, e.g., $I + e_{12} + e_{23}$, from which we can see that if W is a word in which *not* every variable has even exponent-sum, then when evaluated at some set of arguments in G, it does not go to an element of exponent 2; so for such words, $W^2 = 1$ is not an identity of $\mathbf{V}$. It follows that in any *free* group in $\mathbf{V}$, a product of two elements of exponent 2 has exponent 2. (For by the above observation, each of these elements must be represented by a word where all generators have even exponent-sums, hence so will their product.) On the other hand, in G the elements $I + e_{23}$ and $I + e_{12}$ both have order 2, while their product, $I + e_{12} + e_{23}$, has order 4; so G cannot be embedded in a free group in $\mathbf{V}$.

Having gotten parallel results, Theorems 1 and 10, for semigroups and groups, one might expect the corresponding result to hold for monoids. But instead one has

Theorem 11. *A finite monoid M is embeddable in a relatively free monoid if and only if either* (i) *the only invertible element of M is* 1, *or* (ii) *every element of M is invertible. In each case, M is in fact embeddable in a* finite *relatively free monoid.*

Proof. Suppose M is embeddable in a monoid F free on generators $x_1,\dots,x_r$ in a monoid variety $\mathbf{V}$. If (i) does not hold, so that M has an invertible element other than 1, then we get a relation $U(x_1,\dots,x_r)V(x_1,\dots,x_r)=1$ in F, where U and V are nontrivial monoid words (i.e., not the word 1). Mapping F into the free monoid on one generator x by sending all x_i to x, this becomes a relation $x^d=1$ for some $d>0$. Hence $\mathbf{V}$ satisfies the identity $\xi^d=1$, hence in every monoid in $\mathbf{V}$, and in particular, in M, every element a has an inverse, a^{d-1}, proving (ii). This gives the "only if" direction of the first sentence of the theorem.

In proving the converse, together with the final finiteness assertion, we may assume M has more than one element. Suppose first that (ii) holds. Then M, regarded as a finite group, can be embedded by Theorem 10 in a finite relatively free group F. By finiteness of F, any variety of groups in which F is free will satisfy an identity $\xi^d=1$. This allows us to write the group identities of this variety as monoid identities, replacing inverses everywhere by $(d-1)$st powers. Thus, F can be regarded as a relatively free monoid, giving the desired embedding.

To deal with case (i), recall that a map from a finite set to itself that is either one-to-one or onto is both. Looking at the left action of a finite monoid on itself, it is easily deduced that an element of such a monoid which has a one-sided inverse is invertible. Hence if M is a finite monoid satisfying (i), there are no nontrivial solutions in M to the equation $ab=1$, so $M-\{0\}$ is a subsemigroup of M, which we shall call M_0.

We can now apply the construction of Theorem 1 to M_0, getting a semigroup T_0 in which M_0 embeds, and which satisfies identities (3) obtained from the multiplication table of M_0. If we write T for the monoid $T_0\cup\{1\}$, we see that M embeds in T; I claim, moreover, that T still satisfies these same identities. The proof of Theorem 1 gives all instances of these identities except those where X or Y equals 1. The case where $X=Y=1$ is clear; the case where only one of these, say Y, equals 1 subdivides according to whether X has length 1 or >1. In the former case, the desired equations are seen to reduce to $0=0$; the latter behaves like the case in the proof of that theorem where X and Y both had length >1. Given that the finite monoid T satisfies these identities, it follows as in the proof of that theorem that M embeds in the free monoid on two generators in the variety of monoids generated by T. □

One can, of course, also apply the results of the preceding section to get partial positive and negative results on when *infinite* monoids are embeddable in relatively free monoids.

References

[1] S.I. Adyan, Burnside's Problem and Identities in Groups, Nauka, Moscow, 1975 (in Russian). MR 55 #5753.

[2] G.M. Bergman, The Diamond Lemma for Ring Theory, Adv. in Math. 29 (1978) 178–218 (Some errata and updates are collected at http://math.berkeley.edu/∼gbergman/papers/updates/diamond.html.). MR81b:16001.

[3] Trevor Evans, The number of semigroup varieties, Quart. J. Math. Oxford Ser. (2) 19 (1968) 335–336. MR 38#257.

[4] A. Magidin, Dominions in varieties generated by simple groups, Algebra Universalis 48 (2002) 133–143.

[5] M. Morse, G.A. Hedlund, Unending chess, symbolic dynamics and a problem in semigroups, Duke Math. J. 11 (1944) 1–7. MR5, 202e.

[6] V.L. Murskii, Some examples of varieties of semigroups, Mat. Zametki 3 (1968) 663–670 (in Russian). MR 38#256.

[7] S. Oates, Identical relations in groups, J. London Math. Soc. 38 (1963) 71–78. MR 26#5043.

[8] J. Rhodes, B. Steinberg, The q-theory of finite semigroups, in preparation. Viewable at http://www.fc.up.pt/cmup/home/bsteinbg/qtheor.html.